

DataNet/INTEROP: Advances, Collaborative Interactions, and Germinating for the Long Term

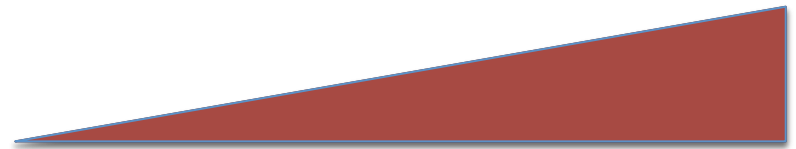
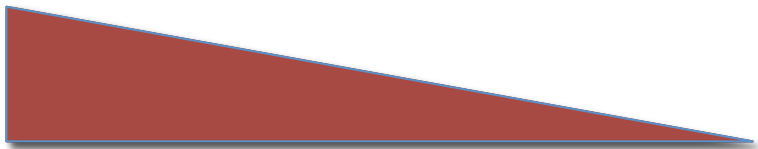
Beth Plale

Professor, School of Informatics and Computing

Director, Data To Insight Center

Managing Director, Pervasive Technology Institute

DataNet/INTEROP Workshop organizer

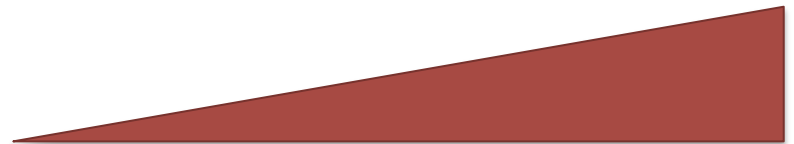
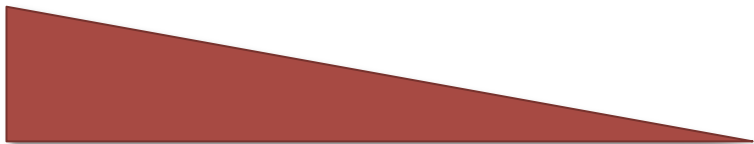


DataNet / INTEROP projects

- NSF DataNet - Sustainable Digital Data Preservation and Access Network Partners - exemplar national and global data research infrastructure organizations (dubbed DataNet Partners) that provide unique opportunities to communities of researchers to advance science and/or engineering research and learning
- NSF Interop - crosscutting program supports community efforts to provide for broad interoperability through the development of mechanisms such as robust data and metadata conventions, ontologies, and taxonomies. Responsible for consensus-building activities and for providing the expertise necessary to turn the consensus into technical standards.

Meeting Format

- The objectives of the workshop will be achieved through a program that draws on notions from the "unconference". The goal of the unconference is to foster emergent ideas or efforts. It can be assumed that data, and for most that means data in the context of informatics or e-Science, scientific data sharing, preservation and interoperability is important to attendees. The unconference, adapted to the project meeting setting, is intended to accomplish the goal of free-form discussion and attainment of objectives.



Workshop Purposes

- I. Project progress from the DataNet and INTEROP projects

Collective expertise in researchers engaged in DataNet and INTEROP is tremendous.

- II. Information exchange between DataNets and INTEROPs

- III. Broader discussion resulting in:

→ Stretch goal of workshop is to plant first seeds for long lived, collective effort that advances scientific sharing and preservation worldwide

Roadmap

- Information exchange
- Interaction between projects: war stories, how to contribute, identifying targets for incorporating results (from INTEROP -> DataNet)
- Creating something greater than sum of parts that aligns with EU partners and beyond

Posters, lightning talks.
This morning.

Birds Of Feather:
informal conversations
on topic of common
interest.

Educational: hearing
from IETF key
founders; European
Union EUDAT and
DAITF efforts

To influence workshop thinking: EarthCube Community Charrette Nov '11

14 Key Capability Categories emerged

Dataset and workflow discovery	Data access services	Tools to Probe, Validate, Verify, and Visualize Data	Data security and trust
Workflow execution management	Data management within workflows	Metadata for workflow and data sets	
Numerical methods and software engineering	Modeling standards and frameworks	Modeling capabilities within Cloud, Grid, HPC, and Science Portals	
Policy enforcement processes	Best practices & Governance Models for definitions & standards	Broad Participation: NGO, industry, domain partners, international	Continuity, Sustainability, & Evolution

Rules of Engagement

- Topics identified as we go; emerge by participant preference.
Suggestions for today's BOF
 - Storage solutions and funding models for long term storage
 - DataCite, its use
 - IETF as model to continue momentum after workshop
 - Issues in using cloud storage for long term availability
 - Dataset discovery and data access services
 - Broadening participation
- Inclusion principle: on stretch goal there are no predestined players. We're trying to build something part grass roots, part top down. It'll take a lot of work by volunteers committed to vision of alliance for advancing scientific preservation. We're looking for committed people to join us.

BOF Session 1 Topics (first day)

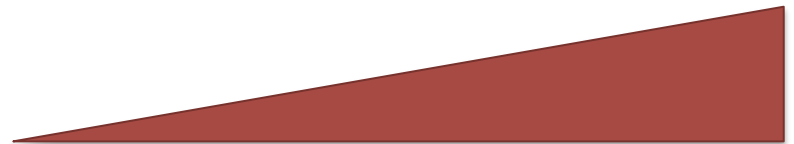
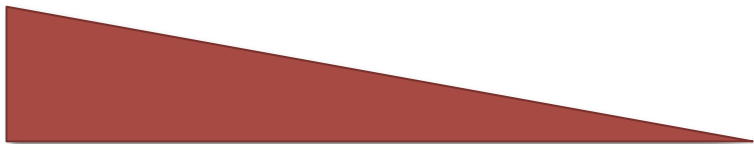
- 1 Storage and funding models; clouds, libraries, repositories
- 2 Metadata and semantics
- 3 Building a lasting coordination group (IETF)
(Beth Plale)
- 4 Data citation
- 5 Education and curriculum for data curation
(Margaret Hedstrom)

BOF Session 2 Topics (day 2)

- ① Data Centric Science Consortium (Beth Plale)
Defining principles for; Includes int'l initiatives
Follow on from BOF 3 on day before
- ② Storage and funding models; clouds, libraries, repositories, cont.
- ③ What will Data Realm look like in 20 yrs?
- ④ Interop between DataNet infrastructures:
what is possible today. (Jim Myers)

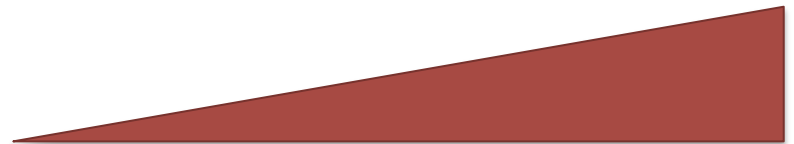
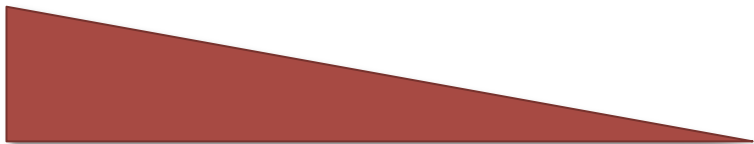
Access By Anyone Anywhere for Advancement of Science and Society

DATA CENTRIC SCIENCE CONSORTIUM



Guiding principles

- **What trying to accomplish:** common needs?
May not be just standards.
- **Who do we want to attract:** librarians, computer scientists, social science, domain scientists. Who do we need to include? Program funding officers, vendors, data center managers
- **Process to accomplish:** we are likely beyond pure “running code” in output.
- Can’t be just national effort. Need to be “Glocal”. IETF keeps stats on where authors come from; it moves meetings to where workers are.



What are we trying to accomplish?

- “Force explicitness and move towards harmonization”
- Define our culture (shared attitudes, values, goals).
- Can we use “rough consensus and running code”?
 - We are to the left of IETF model (see below)
- Open to everyone (non-membership). One represents self as in IETF.
- Top Down or Bottom Up? Like IETF, need both

BRDI model

IETF model

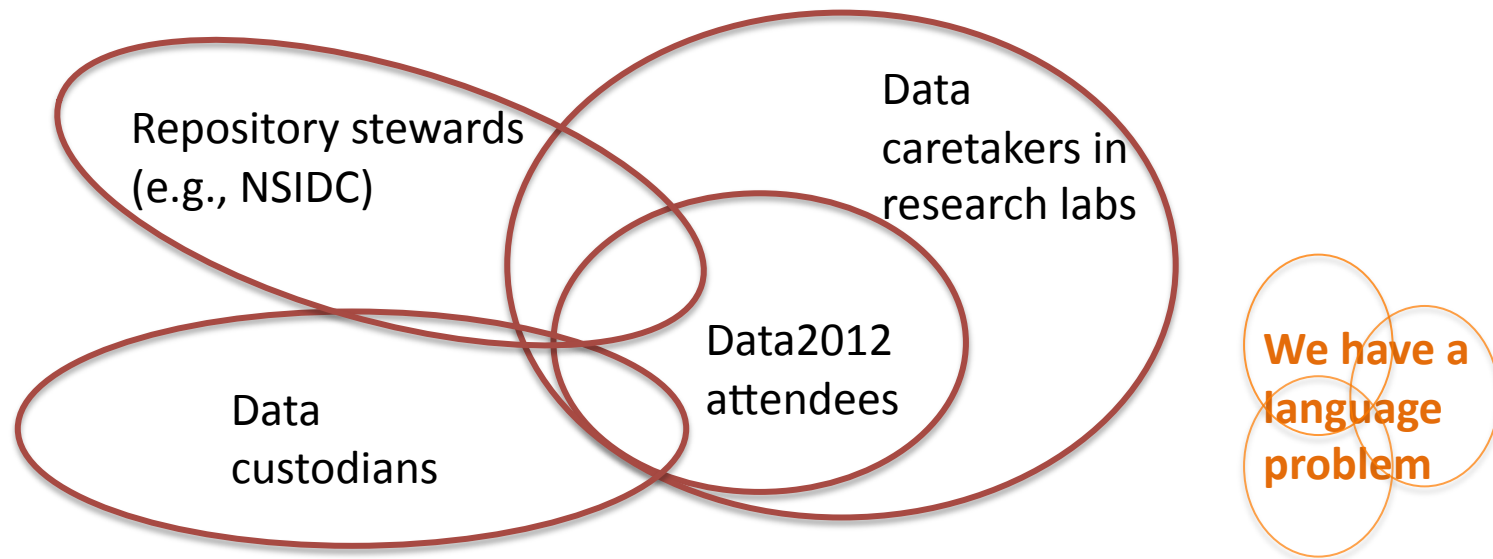
Careful consensus
reports

Rough consensus,
and running code

Access By Anyone Anywhere for Advancement of Science and Society

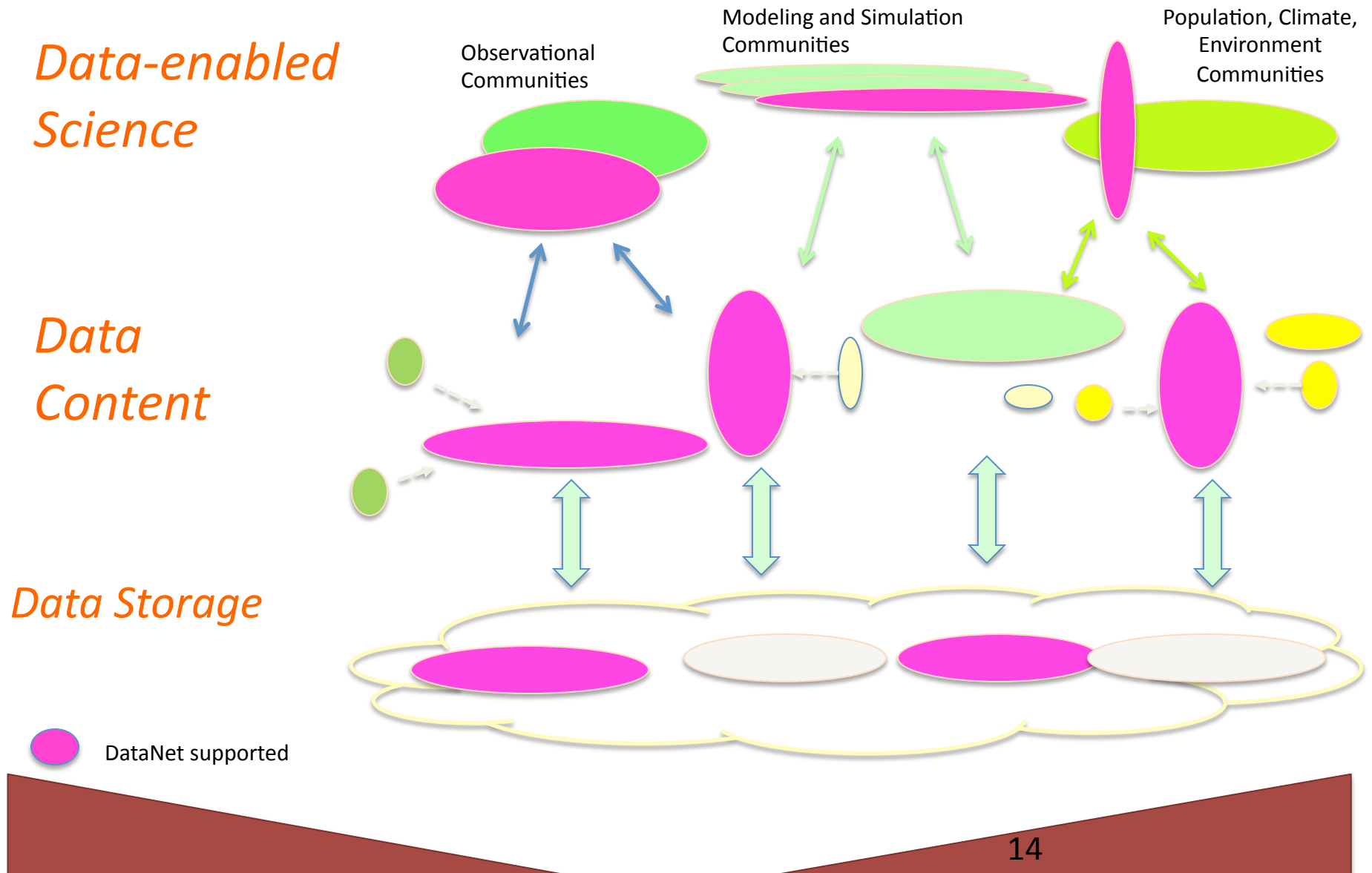
Who do we want to attract?

- See next slide from Robert Pennington. While not used to answer this question, it does frame the discussion by providing a categorization of communities.



Access By Anyone Anywhere for Advancement of Science and Society

DataNet: A Multi-tiered and Multi-Disciplinary Landscape

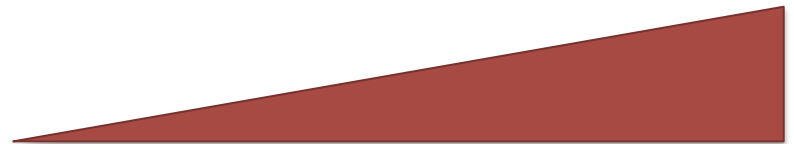
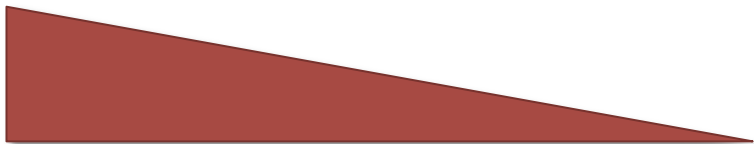


Process to Accomplish

- How big a group to start?
- Work through existing bodies, ESIP, OGC, W3C?
 - Membership organizations have drawbacks.
 - Three possible options with IETF:
 - Graft onto IETF as working groups.
 - Be alternate stream in IETF (includes use rfcs as publishing mechanism).
 - Align with internet society.

First steps

- Organize small group to move forward.
 - Identify topics that are well defined, thorn in side: data to HPC center, DataCite. Where help needed. Likely smaller than “working groups” in ietf sense.
- “Think Glocally”
- DAITF organizing 2 workshops, we organize 2 as well for exchanges.
- Host Challenge “Week”: 2-3 day intensive emersion. Topics from above used in challenge. These could become RFCs.



Data2012 workshop and follow-on materials can be found at:

[http://cgi.cs.indiana.edu/~dsiweb/
data2012/index.php/Main_Page](http://cgi.cs.indiana.edu/~dsiweb/data2012/index.php/Main_Page)

